PDSS homework

The Bayesian network model class has a central position in artificial intelligence:

- As a probabilistic logic knowledge base, it provides a coherent framework to represent beliefs (see Bayesian interpretation of probabilities).
- As a decision network, it provides a coherent framework to represent preferences for actions.
- As a dependency map, it explicitly represents the system of conditional independencies in a given domain.
- As a causal map, it explicitly represents the system of causal relations in a given domain.
- As a decomposable probabilistic graphical model, it parsimoniously represents the quantitative stochastic dependencies (the joint distribution) of a domain and it allows efficient observational inference.
- As an uncertain causal model, it parsimoniously represents the quantitative, stochastic, autonomous mechanisms in a domain and it allows efficient interventional and counterfactual inference.

The goal of the homework is to demonstrate and practise this multifaceted nature of Bayesian networks.

Obligatory and optional subtasks

The minimal level contains the following subtasks (2 points):

1. Select a domain, create candidate variables (5-10), and sketch the structure of the Bayesian network model.
2. Consult it.
3. Quantify the Bayesian networks.
4. Evaluate it with global inference and „information sensitivity of inference" analysis.
5. Extend your Bayesian network into a decision network. Insert an action/decision and a utility node. The utility node preferable depends from the action node and another node semantically related to the action node.
6. Generate a data set from your Bayesian network model.
7. Learn a model from your data.
8. Investigate the structural and parametric differences between the two models.

Optional tasks:

+1. Analyse estimation biases (+1 points).
+2. Investigate the effect of model uncertainty and sample size on learning: vary the strength of dependency in the model (increase underconfidence to decrease information content) and sample size and examine their effect on learning (+2 points).

Consultation

The preliminary approval of your planned homework is mandatory!

Documentation

The homework should be summarized in a document, structured as follows:

Domain description.	10-100 words
Variable definitions, with definitions of their values.	<20 words/variable
Structure of the Bayesian network.	Explain the (preferably) causal order of the variables and interesting independencies in your model. 50-500 words + figure(s).
Quantify the Bayesian networks.	Illustrate your estimation in your model. 50-200 words + table(s)/figure(s).
Evaluate it with global inference[1] and „information sensitivity of inference"[2] analysis.	20-100 words + table(s)/figure(s).
Evaluate the constructed decision network, particularly the actions with maximum expected utilities.	
Compare the structural and parametric differences between the constructed and learnt models.	50-200 words. Check only the existence of edges regardless their orientation.
Analyse estimation biases.	250-500 words + table(s)/figure(s).
Investigate the effect of model uncertainty and sample size on learning.	500-1000 words + table(s)/figure(s).

The overall documentation can be 3-5 pages (minimal) or 5-10 pages (full) long.

Submission

After consultation, the model XML with its documentation should be uploaded to the homework submission site (belated homeworks can be submitted in the first week after the semester, but please try to accomplish it by the 13th week).

Tools

The software system BayesCube with manual is available at

http://redmine.genagrid.eu/projects/bayescubedownload/wiki/Wiki

[1] "Global inference" means that you check the model with inferences using "distant" (e.g. not parent-child) query-evidence pairs.

[2] "information sensitivity of inference analysis" is described in the manual on Section "3.3 Effect of further information on inference". Basically, you can select a single query variable-value, a fixed evidence set and a varying set of evidences sequentially entering into the inference, and the result is visualized in a tree, which shows the sensitivity of the conditional probability of the target for further information.

Hints

Suggested topics: observable everyday activity/scenario, e.g. not finding your mobile, finding an interesting web page, having a good food with friends, hearing an interesting lecture, watching an interesting movie, software developer framework selection, traffic (betweeen home-university), (common sense ;-) weather forecast, mobile phone selection, battery discharge, etc.

Possible biomedical topics: asthma exacerbation, game addiction, flu, melanome, t2diabetes, hypertension, allergy, depression, obesity (effect of diets).

1. Prefer causality, i.e. temporal direction and mechanisms (easier estimation of conditionals).
2. Do not use variables with more values than 5 (binary variables usually suffice).
3. Do not use aggregate, semantic variables (with semantic relations).
4. Save and version your models.

Reference

Russel-Norvig: Artificial intelligence: a modern approach (2nd< edition)

- Chapter 14.,16. (optional chapters 13-16, 18-20)